PEMODELAN REGRESI KUANTIL DENGAN ESTIMATOR *SPLINE* KUBIK MULTIVARIABEL PADA DATA PASIEN PENDERITA DEMAM BERDARAH *DENGUE*

M. RAZY QARAR FAIRUZZABADI
H051201070

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

PEMODELAN REGRESI KUANTIL DENGAN ESTIMATOR *SPLINE* KUBIK MULTIVARIABEL PADA DATA PASIEN PENDERITA DEMAM BERDARAH *DENGUE*

M. RAZY QARAR FAIRUZZABADI
H051201070

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024

SKRIPSI

PEMODELAN REGRESI KUANTIL DENGAN ESTIMATOR *SPLINE* KUBIK MULTIVARIABEL PADA DATA PASIEN PENDERITA DEMAM BERDARAH *DENGUE*

M. RAZY QARAR FAIRUZZABADI
H051201070

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada tanggal 7 Agustus 2024 dan dinyatakan telah memenuhi syarat kelulusan pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan: Pembimbing Tugas Akhir,	Mengetahui: Ketua Program Studi,
Dr. Anna Islamiyati, S.Si., M.Si. NIP. 19770808 200501 2 002	Dr. Anna Islamiyati, S.Si., M.Si. NIP. 19770808 200501 2 002

PERNYATAAN KEASLIAN SKRIPSI
DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Pemodelan Regresi Kuantil Dengan Estimator *Spline* Kubik Multivariabel Pada Data Pasien Penderita Demam Berdarah *Dengue*" adalah benar karya saya dengan arahan dari pembimbing Dr. Anna Islamiyati, S.Si., M.Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbirkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 7 Agustus 2024

10000
METERAI TEMPEL
FC5B1ALX325661333

M. Razy Qarar Fairuzzabadi
NIM H051201070

UCAPAN TERIMA KASIH

Segala puji bagi Allah *Subhanahu Wa ta'ala* atas segala limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "**Pemodelan Regresi Kuantil Dengan Estimator *Spline* Kubik Multivariabel Pada Data Pasien Penderita Demam Berdarah *Dengue***". Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam* beserta keluarga dan para sahabatnya.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Pembimbing Utama yang dengan sabar memberi masukan dan motivasi dalam penulisan skripsi ini. Terima kasih kepada **Ibu Anisa, S.Si., M.Si.** dan **Ibu Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.** selaku Tim Penguji yang senantiasa memberikan saran dan kritikan yang membangun dalam penyempurnaan penulisan skripsi ini. Terima kasih kepada **Pimpinan Universitas Hasanuddin**, **Departemen Statistika**, **Jajaran Dosen**, dan **Staf Departemen Statistika** yang telah membekali ilmu dan kemudahan selama penulis menempuh studi.

Dengan rasa hormat, penulis mengucapkan terima kasih kepada kedua orang tua, Ibunda **Rahmaniah Hasan** (almarhumah), Ayahanda **Nur Ikhwan**, dan Ayahanda **Haikal Hasan** (almarhum) atas inspirasi, pendidikan serta cinta dan kasih sayang yang telah mengiringi setiap langkah penulis dengan doa dan restunya. Terima kasih juga kepada kakak tersayang, **Ryaski Amari Musfirah** atas segala bantuan, motivasi, dan semangat tiada henti yang telah diberikan kepada penulis. Tak lupa, penulis juga berterima kasih kepada Ibu **Anisa** yang telah menggantikan peran seorang ibu bagi penulis dengan memberikan kasih sayang, arahan, dan bimbingan yang sangat berarti. Serta untuk keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini.

Terima kasih kepada teman-teman di **STATISTIKA 2020** atas kebersamaan, semangat, dan motivasi yang telah diberikan selama masa studi di bangku kuliah. Juga, terima kasih kepada keluarga besar Himastat FMIPA Unhas, khususnya **POIS20N** atas ilmu dan pengalaman yang berharga. Penulis bangga menjadi bagian dari keluarga ini dan mengajak untuk tetap **BERSAMA DAN KUAT DI DALAM PERBEDAAN**. Terima kasih kepada kakak-kakak **INTEGRAL 2018**, khususnya Kak **Fadhil Al Anshory** dan Kak **Juni Wahdaniyah** atas seluruh pengalaman, pembelajaran, dan motivasi yang diberikan selama berorganisasi. Kepada teman-teman **GAU22IAN**, terima kasih karena telah menjadi tempat bertukar cerita bagi penulis. Tetaplah **BERSAMA MENGANGKASA**. Dan kepada sahabat-sahabat penulis di **GUARDIAN OF VALUES**, **RUANG DISKUSI**, **INSAP**, dan semua yang tak sempat saya sebutkan, terima kasih atas segala momen kebersamaan, pembelajaran, dan diskusi yang terus membangun penulis.

Penulis menyadari bahwa masih banyak terdapat kekurangan dalam penyusunan skripsi ini. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Penulis,

M. Razy Qarar Fairuzzabadi

ABSTRAK

M. RAZY QARAR FAIRUZZABADI. **Pemodelan Regresi Kuantil Dengan Estimator *Spline* Kubik Multivariabel Pada Data Pasien Penderita Demam Berdarah *Dengue*** (dibimbing oleh Anna Islamiyati).

Latar Belakang. Demam Berdarah *Dengue* (DBD) adalah penyakit menular yang berpotensi menimbulkan wabah dan ditandai dengan penurunan jumlah kadar trombosit pada pasien. Faktor- faktor yang mempengaruhi kadar trombosit pada pasien penderita DBD adalah suhu tubuh dan sel darah putih. Pola Perubahan kadar trombosit pasien penderita DBD berdasarkan suhu tubuh dan sel darah putih menyebabkan data dalam penelitian ini dianalisis menggunakan regresi kuantil dengan estimator *spline* kubik multivariabel. **Tujuan**. Penelitian ini bertujuan untuk mengestimasi fungsi regresi kuantil dengan estimator *spline* kubik multivariabel dan menerapkannya untuk memodelkan data kadar trombosit pasien penderita DBD pada kuantil $\tau = 0.25,\ 0.50,$ dan 0.75. Model dengan nilai *Generelized Cross Validation* (GCV) minimum dipilih sebagai model optimal. **Metode**. Penelitian ini terdiri dari dua tahap, yaitu 1) memperoleh estimasi fungsi regresi kuantil dengan estimator *spline* kubik multivariabel dan 2) memodelkan data kadar trombosit pada pasien penderita DBD dengan regresi kuantil menggunakan estimator *spline* kubik multivariabel. **Hasil**. Model regresi kuantil dengan estimator *spline* kubik multivariabel yang optimal diperoleh dengan menggunakan 3 titik knot untuk kuantil $\tau = 0.50$, yang menghasilkan nilai GCV minimum sebesar 364.874. **Kesimpulan**. Model regresi kuantil dengan estimator *spline* kubik multivariabel memberikan hasil estimasi yang menunjukkan pola perubahan kadar trombosit pada pasien penderita DBD, yang dipengaruhi oleh suhu tubuh dan jumlah sel darah putih.

Kata Kunci: Demam Berdarah *Dengue*, Kadar Trombosit, Regresi Kuantil, *Spline*, *Generelized Cross Validation*

ABSTRACT

M. RAZY QARAR FAIRUZZABADI. **Quantile Regression Modeling with Cubic Spline Estimator Multivariable Data on Patients with Dengue Fever** (supervised by Anna Islamiyati).

Background. Dengue fever (DHF) is an infectious disease that has the potential to cause an outbreak and is characterized by a decrease in the number of platelet levels in patients. Factors that affect platelet levels in patients with DHF are body temperature and white blood cells. Changes in platelet levels of patients with DHF based on body temperature and white blood cells cause the data in this study to be analyzed using quantile regression with a multivariable cubic spline estimator. **Objective**. This study aims to estimate the quantile regression function with a multivariable cubic spline estimator and apply it to model platelet level data of patients with DHF at quantiles $\tau = 0.25,\ 0.50,$ dan 0.75. The model with the minimum Generelized Cross Validation (GCV) value was selected as the optimal model. **Method**. This research consists of two stages, namely 1) obtaining the estimation of quantile regression function with multivariable cubic spline estimator and 2) modeling platelet level data in patients with DHF with quantile regression using multivariable cubic spline estimator. **Results**. The optimal quantile regression model with multivariable cubic spline estimator was obtained using 3 knot points for quantile $\tau = 0.50$, which resulted in a minimum GCV value of 364.874. **Conclusion**. The quantile regression model with multivariable cubic spline estimator provides estimation results that show the pattern of changes in platelet levels in patients with dengue fever, which is influenced by body temperature and white blood cell count.

Keywords: Dengue Hemorrhagic Fever, Thrombocyte Level, Quantile Regression, *Spline*, *Generelized Cross Validation*

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
Regresi	Analisis statistik yang digunakan untuk mengidentifikasi hubungan antara variabel respon berdasarkan satu atau lebih variabel prediktor
Kuantil	Nilai yang membagi data menjadi bagian-bagian yang sama
Estimator	Metode yang digunakan untuk memperkirakan nilai parameter populasi berdasarkan data sampel
Spline	Fungsi matematika yang digunakan untuk membuat model data yang fleksibel dan dapat menangani hubungan yang kompleks antara variabel
Multivariabel	Analisis yang melibatkan lebih dari dua variabel secara bersamaan
Estimasi	Proses memperkirakan nilai dari parameter populasi berdasarkan data sampel
Parameter	Nilai yang menggambarkan sifat atau karakteristik suatu populasi
Pencilan	Nilai yang sangat berbeda dari sebagian besar data lainnya dalam satu set data
Distribusi	Pola penyebaran data atau probabilitas kejadian dalam ruang sampel
Parametrik	Metode yang mengasumsikan bahwa data mengikuti distribusi tertentu
Nonparametrik	Metode yang tidak mengasumsikan bentuk distribusi data
Normalitas	Distribusi data yang mengikuti pola distribusi normal
Heterogenitas	Variasi atau perbedaan yang ada dalam suatu kumpulan data atau populasi
Least Absolute Deviation	Metode estimasi parameter model regresi yang meminimumkan jumlah absolut deviasi
Tersegmen	Pembagian fungsi menjadi segmen-segmen yang masing-masing diwakili oleh polinomial berbeda
Titik knot	Titik-titik tertentu pada domain variabel independen di mana bentuk fungsi *spline* berubah

DAFTAR LAMBANG DAN SINGKATAN

Lambang/singkatan	Arti dan Penjelasan
y_i	Variabel respon pada pengamatan ke-i
x_i	Variabel prediktor pada pengamatan ke-i
ε_i	*Error* ke-i
x_{ij}	Pengamatan ke-i pada variabel prediktor ke-j
τ	Nilai kuantil
$y_i(\tau)$	Variabel respon ke-i pada kuantil ke-θ
$\beta(\tau)$	Penduga parameter pada kuantil ke-θ
$\varepsilon_i(\tau)$	*Error ke-*i pada kuantil ke-θ
$\boldsymbol{y}$	Vektor kolom berukuran $n \times 1$ dari variabel respon
$\boldsymbol{X}$	Matriks berukuran $n \times (k+1)$ dari variabel prediktor
$\boldsymbol{\beta}$	Vektor kolom berukuran parameter regresi
$\boldsymbol{\varepsilon}$	Vektor kolom berukuran $n \times 1$ dari *error*
$\rho_\theta(\varepsilon)$	*Loss function* ke-θ dari ε
$E[\rho_\theta(\varepsilon)]$	Ekspektasi *error ke-*i pada kuantil ke-θ
$Q_\theta(y\|x)$	Fungsi kuantil ke-θ dari variabel y dengan syarat x
$\hat{\beta}(\tau)$	Penduga dar parameter regresi pada kuantil ke-θ
c_j	Nilai koefisien fungsi tujuan
c_b	Nilai koefisien dari variabel yang menjadi basis
v_b	Nama-nama dari variabel yang menjadi basis
w_b	Nilai ruas kanan dari kendala
z_j	Nilai koefisien fungsi objektif
a_{ij}	Nilai koefisien kolom kunci
q	Orde polinomial *spline truncated*
$(x_i - K_h)_+^q$	Fungsi *polynomial truncated*
K_h	Nilai titik knot ke-h
$f(x_{ij})$	Fungsi regresi dari pengamatan ke-i pada variabel prediktor ke-j
β_{lj}	Parameter polinomial pada orde ke-l dan prediktor ke-j
$\beta_{(q+h)j}$	Parameter *truncated* pada titik knot ke-$(q+h)$ dan prediktor ke-j
K_{hj}	Nilai titik knot ke-h pada prediktor ke-j
$\boldsymbol{X}[K]$	Matriks berukuran $n \times (1+q+qr)$ dari model yang bergantung pada titik knot
OLS	*Ordinary Least Square*
BLUE	*Best Linear Unbiased Estimator*
GCV	*Generelized Cross Validation*
LAD	*Least Absolute Deviation*
OBD	Operasi Baris Dasar

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Penyakit Demam Berdarah *Dengue* (DBD) adalah penyakit menular yang berpotensi menimbulkan wabah di suatu wilayah. Di Indonesia, penyakit ini menjadi salah satu masalah kesehatan yang perlu diwaspadai karena perkembangan penyakitnya yang pesat dan berisiko kematian dalam waktu singkat. Penyakit ini disebabkan oleh virus *dengue* yang masuk ke tubuh manusia melalui gigitan nyamuk *Ae Aegypti* (Amelia dkk., 2020). Penurunan jumlah kadar trombosit pada pasien menjadi salah satu indikator terjangkitnya DBD (Rahayuningrum & Morika, 2019). Kriteria laboratorium *World Health Organization* (WHO) mengenai jumlah trombosit yang rendah (*trombositopenia*) dan kebocoran plasma yang ditandai dengan *hemokonsentrasi* merupakan indikator keparahan penyakit DBD. *Trombositopenia* biasanya terjadi antara hari ketiga hingga kedelapan setelah sakit, sebelum terdapat peningkatan hematokrit dan penurunan suhu tubuh. Selain jumlah trombosit dan suhu tubuh, pemeriksaan medis lain seperti pemeriksaan sel darah putih juga diperlukan (Hadijah, 2020).

Data yang memuat variabel respon dan prediktor umumnya dianalisis menggunakan pendekatan regresi. Analisis regresi merupakan salah satu teknik analisis statistik yang digunakan untuk melihat pola hubungan dan pengaruh antara dua variabel, yaitu variabel respon berdasarkan satu atau lebih variabel prediktor melalui estimasi kurva regresi (Puteri dkk., 2020). Salah satu metode yang umum digunakan untuk mengestimasi parameter dalam persamaan regresi adalah *Ordinary Least Square* (OLS). Prinsip metode ini adalah meminimumkan jumlah kuadrat *error* untuk mendapatkan estimator terbaik. Penggunaan metode OLS membutuhkan beberapa asumsi klasik yang harus dipenuhi, seperti normalitas, homoskedastisitas, dan autokorelasi. Apabila asumsi terpenuhi, maka hasil estimasi parameter akan menjadi *Best Linear Unbiased Estimator* (BLUE). Namun, asumsi ini seringkali tidak terpenuhi, terutama ketika data mengandung pencilan. Pencilan dapat menyebabkan bentuk data tidak simetris, sehingga nilai *mean* menjadi sensitif terhadap kehadiran pencilan. Selain itu, pencilan juga dapat menyebabkan bias dalam hasil estimasi persamaan regresi. Oleh karena itu, salah satu teknik analisis regresi yang dapat digunakan untuk mengatasi data yang mengandung pencilan adalah dengan menggunakan regresi kuantil (Wahyudi & Zain, 2014).

Koenker (2005) memperkenalkan regresi kuantil, yaitu metode analisis regresi yang digunakan untuk membagi data yang diduga memiliki perbedaan nilai hasil estimasi pada kuantil tertentu. Pendekatan ini melibatkan pendugaan fungsi kuantil dari sebaran bersyarat respon pada berbagai nilai kuantil yang diinginkan. Setiap kuantil menggambarkan titik tertentu pada bentuk distribusi data. Regresi kuantil memiliki beberapa keunggulan dibandingkan metode OLS, salah satunya adalah fleksibilitas dalam memodelkan data dengan distribusi bersyarat heterogen (Mahmuda dkk., 2015). Keunggulan lainnya adalah metode ini tidak terpengaruh oleh pencilan, sehingga tidak mengganggu stabilitas data. Regresi kuantil dapat diterapkan pada berbagai kasus regresi, baik parametrik maupun nonparametrik (Wahyudi & Zain, 2014).

Regresi parametrik adalah metode regresi yang diasumsikan bentuk kurva regresinya sudah diketahui, seperti linier, kuadratik, atau kubik. Namun, tidak semua data dapat dianalisis dengan pendekatan parametrik karena keterbatasan informasi mengenai bentuk dan pola hubungan antara variabel respon dan variabel prediktor. Oleh karena itu, regresi nonparametrik hadir sebagai alternatif, di mana bentuk dan pola hubungan antara variabel respon dan prediktor tidak diketahui (Hidayat dkk., 2017). Dalam perkembangannya, berbagai estimator telah dikembangkan untuk regresi nonparametrik pada berbagai jenis data aplikasi, salah satunya adalah *spline*. *Spline* merupakan salah satu estimator regresi nonparametrik yang telah berkembang dengan sangat pesat karena memiliki tingkat fleksibilitas tinggi, di mana model akan secara otomatis mencari pola yang sesuai dengan data. Beberapa jenis estimator *spline* yang telah dikembangkan oleh para peneliti, di antaranya *spline truncated*, *smoothing spline*, dan *penalized spline* (Anisa dkk., 2023).

Spline terdiri dari potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Fungsi polinomial pada *spline* memiliki sifat yang lebih khusus dan fleksibel dibandingkan fungsi polinomial pada umumnya. *Spline* juga memiliki titik penghubung yang disebut titik knot (Nurdiani dkk., 2017). Keunggulan *spline* terletak pada kemampuannya mengatasi bentuk pola data yang cenderung naik dan turun secara tajam. Model *spline* yang optimal dapat diperoleh melalui nilai *Generelized Cross Validation* (GCV) minimum (Utami, 2018). Penelitian mengenai *spline* telah dikembangkan dalam konteks regresi kuantil dengan satu variabel prediktor, di antaranya Faeni (2021) mengestimasi parameter regresi kuantil menggunakan *smoothing spline* pada data Indeks Kualitas Udara (IKU) di Indonesia, sementara Siregar dkk. (2022) mengestimasi parameter regresi kuantil menggunakan *penalized spline* pada data harga minyak mentah dunia selama pandemi *Covid-19*. Di sisi lain, penelitian mengenai regresi kuantil juga telah diperluas dengan mempertimbangkan pola kuadratik. Sebagai contoh, Anisa dkk. (2023) mengestimasi parameter regresi kuantil menggunakan *spline* kuadratik untuk menganalisis pola perubahan trombosit terhadap hematokrit pada pasien penderita DBD. Namun, penelitian-penelitian tersebut tidak mempertimbangkan penggunaan lebih dari satu variabel prediktor untuk menjelaskan variabel respon dan hanya mengeksplorasi model dengan memperhatikan pola linier dan kuadratik. Oleh karena itu, penelitian ini akan mengkaji penggunaan regresi kuantil *spline* dalam bentuk multivariabel dengan melibatkan beberapa variabel prediktor serta mempertimbangkan penggunaan *spline* dengan memperhatikan pola yang lebih kompleks, khususnya pola kubik. Penggunaan *spline* dengan pola yang lebih kompleks dapat memungkinkan penyesuaian yang lebih baik terhadap pola kompleks dan perubahan tajam dalam data serta menghasilkan kurva yang lebih halus dan akurat. Model ini akan diaplikasikan pada data kadar trombosit pasien penderita DBD berdasarkan faktor suhu tubuh dan sel darah putih pasien. Pola perubahan naik-turunnya kadar trombosit menyebabkan data dapat dianalisis menggunakan regresi kuantil dengan estimator *spline* kubik multivariabel.

1.2 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, batasan masalah pada penelitian ini adalah sebagai berikut.

1. Jenis estimator *spline* yang digunakan adalah *spline truncated*.
2. Nilai kuantil yang digunakan terdiri dari $\tau = 0.25$, 0.50, dan 0.75.
3. Jumlah titik knot yang diuji berkisar antara satu hingga tiga titik knot.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Memperoleh hasil estimasi fungsi regresi kuantil dengan estimator *spline* kubik multivariabel.
2. Memperoleh model data kadar trombosit pada pasien penderita DBD menggunakan regresi kuantil dengan estimator *spline* kubik multivariabel.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Menambah wawasan keilmuan dan pengetahuan dalam mengimplementasikan estimasi model regresi kuantil dengan estimator *spline* kubik multivariabel pada data kadar trombosit pasien penderita DBD.
2. Memberikan informasi yang akurat dan lebih terperinci mengenai kadar trombosit berdasarkan suhu tubuh dan sel darah putih pada pasien penderita DBD, sehingga dapat menjadi acuan bagi akademisi maupun praktisi dalam penanganan pasien penderita DBD.

1.5 Landasan Teori

1.5.1 Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu metode yang digunakan untuk mengestimasi kurva regresi. Kurva dalam regresi nonparametrik tidak diasumsikan secara spesifik, melainkan hanya diasumsikan mulus (*smooth*) dan termuat dalam suatu fungsi tertentu. Hal ini memberikan tingkat fleksibilitas yang tinggi pada regresi nonparametrik. Selain itu, estimasi kurva dengan pendekatan ini dapat disesuaikan dengan data tanpa dipengaruhi oleh subjektivitas peneliti (Mariati dkk., 2019). Model regresi nonparametrik secara umum dapat dituliskan sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i, i = 1,2, \dots, n \tag{1}$$

dengan y_i sebagai variabel respon pada pengamatan ke-i, $f(x_i)$ adalah persamaan kurva regresi yang belum diketahui dengan x_i merupakan variabel prediktor pada pengamatan ke-i dan ε_i merupakan *error* dengan nilai *mean* sebesar 0 dan variansi sebesar σ^2 (Eubank, 1999).

1.5.2 Regresi Kuantil

Koenker (2005) memperkenalkan metode regresi kuantil pada tahun 1978. Regresi kuantil merupakan salah satu metode dalam analisis regresi yang digunakan untuk menduga hubungan antara variabel respon dengan variabel prediktor yang tidak hanya berfokus pada ukuran pemusatan dari variabel respon, tetapi pada berbagai kuantil (Puteri dkk., 2020). Metode ini berfungsi saat data tidak berdistribusi homogen (*heterogenous*) atau tidak berbentuk standar, seperti ketidaksimetrisan atau ketidaknormalan distribusi serta terdapat ekor pada sebaran data atau *truncated distribution* (Santri & Hanike, 2020). Selain itu, metode ini tidak dipengaruhi oleh adanya pencilan dalam data, sehingga pencilan tidak mengganggu kestabilan dari data yang diperoleh (Idris dkk., 2018).

Misalkan diberikan data $\{x_{i1}, x_{i2}, \ldots, x_{ik}, y_i\}$ dengan $i = 1,2, \ldots, n$ dan $j = 1,2, \ldots, k$ adalah himpunan berpasangan dari variabel acak yang berdistribusi secara independen dan tidak identik dengan kuantil $\tau \in (0,1)$. Data ini memiliki fungsi sebaran peluang bersyarat $F(y|x_i) = P(Y \leq y|x_i)$ dan fungsi invers $F^{-1}(\tau) = \inf\{y: F(y) \geq \tau\}$ yang merupakan kuantil ke-τ dari variabel respon y (Davino dkk., 2013). Secara umum, model regresi kuantil linier khusus untuk kuantil bersyarat dari variabel respon y_i adalah sebagai berikut.

$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_j(\tau)x_{ij} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau) \tag{2}$$

dengan $i = 1,2, \ldots, n$ dan $j = 1,2, \ldots, k$

$y_i(\tau)$	:	variabel respon ke-i pada kuantil ke-τ
β_0	:	koefisien konstanta atau intersep pada sumbu y
τ	:	nilai kuantil
$\beta(\tau)$	:	penduga parameter pada kuantil ke-τ
x_{ij}	:	pengamatan ke-i pada variabel prediktor ke-j
$\varepsilon_i(\tau)$	:	*error* ke-i pada kuantil ke-τ

Apabila model regresi kuantil disajikan dalam bentuk matriks, Persamaan (2) dapat dituliskan sebagai berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_k(\tau) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \tag{3}$$

Kemudian Persamaan (3) dapat ditulis dalam bentuk model linier sebagai berikut.

$$\boldsymbol{y}(\tau) = \boldsymbol{X\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \tag{4}$$

dengan

$\boldsymbol{y}(\tau)$	:	vektor kolom berukuran $n \times 1$ dari variabel respon y
$\boldsymbol{X}$	:	matriks berukuran $n \times (k+1)$ dengan baris n merupakan observasi pada kolom k variabel prediktor ke-j dengan $j = 1,2, \ldots k$
$\boldsymbol{\beta}$	:	vektor kolom berukuran $(k+1) \times 1$ dari parameter β_j dengan $j = 1,2, \ldots k$
$\boldsymbol{\varepsilon}$	:	vektor kolom berukuran $n \times 1$ dari *error* ε_i

Menurut Koenker (2005), kuantil dapat dicari dengan menggunakan optimasi, yaitu dengan mendefinisikan *loss function*.

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & , \varepsilon \geq 0 \\ (1-\tau)\varepsilon & , \varepsilon < 0 \end{cases} \tag{5}$$

dengan $\rho_\tau(\varepsilon)$ merupakan *loss function* ke-τ dari ε, di mana ε merupakan *error*, yaitu $y - \hat{y}$ dan τ merupakan konstanta dengan nilai $0 < \tau < 1$.

Loss function adalah fungsi asimetris, kecuali pada $\tau = \frac{1}{2}$. Sedangkan, untuk nilai τ lainnya, *loss function* memberikan bobot sebesar $(1-\tau)$ untuk *error* negatif dan bobot sebesar τ untuk *error* positif. Hal ini dilakukan agar kuantil yang diperoleh sesuai dengan nilai τ yang ditetapkan. Kuantil ke-τ dari F_y dapat diperoleh dengan meminimumkan ekspektasi *loss* pada Persamaan (6). Dengan demikian, ekspektasi *loss* dengan *error* $\varepsilon = y - \hat{y}$ adalah sebagai berikut.

$$\begin{aligned} E[\rho_\tau(\varepsilon)] &= E[\rho_\tau(y-\hat{y})] \\ &= \int_{-\infty}^{\infty} \rho_\tau(y-\hat{y}) f(y)dy \end{aligned}$$

Perbedaan *error* membuat ekspektasi *loss* dibagi menjadi dua bagian, yaitu sebagai berikut.

$$E[\rho_\tau(y-\hat{y})] = \int_{-\infty}^{\hat{y}} (1-\tau)(y-\hat{y})f(y)dy + \int_{\hat{y}}^{\infty} \tau\,(y-\hat{y})f(y)dy \tag{6}$$

Selanjutnya, Persamaan (6) diminimumkan terhadap $\hat{y}$

$$\begin{aligned}
\frac{\partial}{\partial\hat{y}} E[\rho_\tau(y-\hat{y})] &= 0 \\
\frac{\partial}{\partial\hat{y}}\left[\int_{-\infty}^{\hat{y}} (1-\tau)(y-\hat{y})f(y)dy + \int_{\hat{y}}^{\infty} \tau\,(y-\hat{y})f(y)dy\right] &= 0 \\
\frac{\partial}{\partial\hat{y}}\left[(1-\tau)\int_{-\infty}^{\hat{y}} (y-\hat{y})f(y)dy + \tau\int_{\hat{y}}^{\infty} (y-\hat{y})f(y)dy\right] &= 0 \\
(1-\tau)\left[(y-\hat{y})f(y)|_{-\infty}^{\hat{y}} + \int_{-\infty}^{\hat{y}} \frac{\partial}{\partial\hat{y}}(y-\hat{y})f(y)dy\right] \\
+\tau\left[(y-\hat{y})f(y)|_{\hat{y}}^{\infty} + \int_{\hat{y}}^{\infty} \frac{\partial}{\partial\hat{y}}(y-\hat{y})f(y)dy\right] &= 0 \\
(1-\tau)\left[(y-\hat{y})f(y)|_{y=\hat{y}} + \int_{-\infty}^{\hat{y}} f(y)dy\right] + \tau\left[(y-\hat{y})f(y)|_{y=\hat{y}} + \int_{\hat{y}}^{\infty} f(y)dy\right] &= 0 \\
(1-\tau)[0 + F_y(\hat{y})] + \tau[0 - (1 - F_y(\hat{y}))] &= 0 \\
(1-\tau)F_y(\hat{y}) - \tau\left(1 - F_y(\hat{y})\right) &= 0 \\
(1-\tau)F_y(\hat{y}) - \tau + \tau F_y(\hat{y}) &= 0 \\
F_y(\hat{y}) - \tau &= 0
\end{aligned}$$

diperoleh

$$F_y(\hat{y}) = \tau \tag{7}$$

Sehingga kuantil ke-τ adalah solusi dari F_y (Davino dkk., 2013).

1.5.3 Estimasi Parameter Regresi Kuantil

Terdapat beberapa metode yang dapat digunakan untuk mengestimasi parameter dalam persamaan regresi. Salah satu pendekatan standar dalam penentuan model regresi klasik dan estimasi parameternya adalah metode kuadrat terkecil atau *Ordinary Least Square* (OLS), di mana prinsip utama metode ini adalah meminimumkan jumlah kuadrat *error*. Namun, metode OLS sangat rentan dengan adanya pencilan dikarenakan dapat menyebabkan hasil dari estimasi parameter menjadi tidak stabil. Selain itu, metode OLS hanya dapat digunakan untuk memberikan solusi terkait permasalahan *mean*. Oleh karena itu, Koenker pada tahun 1978 mengembangkan metode alternatif, yaitu regresi kuantil. Dalam regresi kuantil, parameter diestimasi dengan meminimumkan jumlah *absolute error* yang lebih dikenal dengan *Least Absolute Deviation* (LAD) (Amyad dkk., 2018).

Pada regresi kuantil, *error* diberikan bobot yang berbeda. Bobot yang digunakan, yaitu τ untuk nilai *error* yang lebih besar atau sama dengan nol dan $1-\tau$ untuk nilai *error* yang kurang dari nol. Perkalian antara *error* dengan bobot yang diberikan disebut sebagai *loss function* (ρ_τ) yang dinyatakan sebagai berikut (Balami, 2017).

$$\rho_\tau(\varepsilon) = \sum_{i=1,\varepsilon_i\geq 0}^{n} \tau|\varepsilon_i| + \sum_{i=1,\varepsilon_i<0}^{n} (1-\tau)|\varepsilon_i| \tag{8}$$

Dengan demikian, pada regresi kuantil terdapat fungsi kuantil ke-τ dari variabel y dengan syarat x mempertimbangkan penduga $\beta(\tau)$, sehingga diperoleh solusi untuk permasalahan tersebut dinyatakan sebagai berikut.

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau\big(y_i - Q_\tau(y|x)\big) \tag{9}$$

dengan $\rho_\tau(\varepsilon)$ sebagai *loss function*, di mana τ adalah indeks kuantil dengan $\tau \in (0,1)$ dan $Q_\tau(y|x)$ adalah fungsi kuantil ke-τ dari variabel y dengan syarat x. Fungsi untuk kuantil $Q_\tau(y|x)$ dinyatakan sebagai berikut.

$$Q_\tau(y|x) = \boldsymbol{X}_i^T\boldsymbol{\beta}(\tau) \tag{10}$$

Dalam regresi kuantil, pada kuantil ke-τ dari F_y meminimumkan *loss function* dari Persamaan (9) sebagai berikut.

$$\hat{\beta}(\tau) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau\Big(y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}(\tau)\Big) \tag{11}$$

dengan $\rho_\tau(\varepsilon)$ pada Persamaan (11) didefinisikan sebagai berikut.

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon &, \varepsilon \geq 0 \\ (1-\tau)\varepsilon &, \varepsilon < 0 \end{cases}$$

Selanjutnya, akan dibuktikan bahwa *loss function* berbentuk asimetris dengan penjelasan sebagai berikut.

$$\rho_\tau = \big[\tau I(\varepsilon \geq 0) + (1-\tau)I\big((\varepsilon < 0)\big)\big]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon \tag{12}$$

dengan

$$I(\varepsilon \geq 0) = \begin{cases} 1, \varepsilon \geq 0 \\ 0, \varepsilon < 0 \end{cases}$$

dan

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & ,\varepsilon \geq 0 \\ (1-\tau)\varepsilon & ,\varepsilon < 0 \end{cases}$$

Sehingga dapat dibuktikan sebagai berikut.

1. Untuk $\varepsilon \geq 0$

$$\begin{aligned} \rho_\tau &= [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]|\varepsilon| \\ &= [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]\varepsilon \\ &= [\tau 1 + (1-\tau)I(\varepsilon < 0)]\varepsilon \\ &= [\tau + I(\varepsilon < 0) - \tau I(\varepsilon < 0)]\varepsilon \\ &= [\tau + (1 - I(\varepsilon \geq 0)) - \tau(1 - I(\varepsilon \geq 0))]\varepsilon \\ &= [\tau + (1-1) - \tau(1-1)]\varepsilon \\ &= \tau\varepsilon \end{aligned}$$

atau

$$\begin{aligned} \rho_\tau &= [\tau - I(\varepsilon < 0)]\varepsilon \\ &= \left[\tau - \left(1 - I(\varepsilon \geq 0)\right)\right]\varepsilon \\ &= [\tau - (1-1)]\varepsilon \\ &= \tau\varepsilon \end{aligned}$$

2. Untuk $\varepsilon < 0$

$$\begin{aligned} \rho_\tau &= [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]|\varepsilon| \\ &= \left[\tau I(\varepsilon \geq 0) + (1-\tau)I\left((\varepsilon < 0)\right)\right](-\varepsilon) \\ &= [\tau 0 + (1-\tau)I(\varepsilon < 0)](-\varepsilon) \\ &= [(\tau - 1)I(\varepsilon < 0)]\varepsilon \\ &= \left[(\tau-1)\left(1 - I(\varepsilon \geq 0)\right)\right]\varepsilon \\ &= [(\tau-1)(1-0)]\varepsilon \\ &= (\tau-1)\varepsilon \end{aligned}$$

atau

$$\begin{aligned} \rho_\tau &= [\tau - I(\varepsilon < 0)]\varepsilon \\ &= \left[\tau - \left(1 - I(\varepsilon \geq 0)\right)\right]\varepsilon \\ &= [\tau - (1-0)]\varepsilon \\ &= (\tau-1)\varepsilon \end{aligned}$$

Sehingga menjadi

$$\rho_\tau = [\tau I(\varepsilon \geq 0) + (1-\tau)I((\varepsilon < 0))]|\varepsilon| = [\tau - I(\varepsilon < 0)]\varepsilon, \forall_\varepsilon$$

Apabila y merupakan fungsi x yang diketahui dan memiliki fungsi probabilitas $F_{y|x}(y)$, maka kuantil ke-τ dari fungsi tersebut dapat dituliskan sebagai berikut.

$$\min_\beta \tau \int_{i=1;\varepsilon_i \geq 0}^{n} |\varepsilon_i| dF_y(y) + (1-\tau) \int_{i=1;\varepsilon_i < 0}^{n} |\varepsilon_i| dF_y(y) \tag{13}$$

dengan mempertimbangkan $\hat{\beta}(\tau)$, maka diperoleh solusi untuk permasalahan yang dinyatakan pada Persamaan (13) sebagai berikut.

$$\hat{\beta}(\tau) = \underset{\beta \in R^p}{min} \left\{ \tau \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)| + (1-\tau) \sum_{i=1,\varepsilon_i < 0}^{n} |y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}(\tau)| \right\} \quad (14)$$

Solusi dari Persamaan (14) tidak dapat diperoleh secara analitik. Solusi yang dapat digunakan untuk menyelesaikan Persamaan (14) adalah secara numerik, yaitu dengan algoritma simpleks. Algoritma simpleks adalah metode yang dikembangkan oleh Barrodale dan Robert pada tahun 1974, di mana metode ini dapat memberikan solusi permasalahan program linier yang melibatkan beberapa variabel keputusan dengan bantuan komputasi (Davino dkk., 2013). Adapun beberapa istilah yang terdapat dalam algoritma simpleks adalah sebagai berikut.

1. Variabel *Slack*
   Variabel *slack* berfungsi untuk menampung sisa kapasitas pada kendala yang berupa pembatas.
2. Variabel *Surplus*
   Variabel *surplus* berfungsi untuk menampung kelebihan nilai ruas kiri pada kendala yang berupa syarat. Pada kasus regresi kuantil dengan menggunakan metode LAD, variabel *surplus* adalah deviasi bawah yang diboboti dengan $(1-\tau)$.
3. Variabel *Artificial*
   Variabel *artificial* adalah variabel positif yang berfungsi untuk memulai penyelesaian dan harus dijadikan nol pada solusi akhir. Variabel ini digunakan untuk setiap persamaan yang tidak memiliki basis. Pada kasus regresi kuantil dengan menggunakan metode LAD, variabel *artificial* adalah deviasi atas yang diboboti dengan τ.
4. Variabel Basis dan Nonbasis
   Variabel basis dan nonbasis merupakan dua terminologi penting yang akan selalu digunakan di dalam algoritma simpleks. Variabel basis adalah variabel yang bernilai positif dan variabel nonbasis adalah variabel yang bernilai nol.

Algoritma simpleks membutuhkan sebuah tabel atau yang biasa dikenal dengan tabulasi simpleks seperti pada Tabel 1.

Tabel 1. Tabel Metode Simpleks untuk Kasus Regresi Kuantil

c_j			0	0	...	0	τ	...	τ	$(1-\tau)$	...	$(1-\tau)$
c_b	v_b	w_b	x_1	x_2	...	x_n	d_{11}	...	d_{1n}	d_{21}	...	d_{2n}
d_{11}^+	x_1	b_1	a_{ij}									
d_{21}^+	x_2	b_2										
⋮	⋮	⋮										
d_{n1}^+	x_n	b_n										
z_j												
$c_j - z_j$												

Pengisian Tabel 1 diuraikan sebagai berikut.

a. Baris c_j diisi dengan koefisien fungsi tujuan.
b. Kolom c_b diisi dengan koefisien variabel yang menjadi basis.

c. Kolom v_b diisi dengan nama-nama variabel yang menjadi basis (variabel yang menyusun matriks identitas), dalam hal ini diisi dengan variabel *artificial* yang merupakan deviasi atas.

d. Kolom w_b diisi dengan nilai ruas kanan dari kendala.

e. Baris z_j diisi dengan rumus $z_j = \sum d_i a_{ij}$, $j = 1,2, \dots, n$.

Berikut prosedur algoritma simpleks.

1. Mengubah masalah optimasi linier ke dalam bentuk standar dan fungsi tujuan serta kendala-kendala ke dalam bentuk persamaan, yaitu dengan menambahkan variabel *slack*, *surplus*, dan *artificial* terhadap kendala yang berbentuk pertidaksamaan.
2. Menentukan kolom kunci (variabel keluar). Untuk masalah maksimum, memilih nilai $c_j - z_j$ terbesar. Sedangkan untuk masalah minimum, memilih nilai $c_j - z_j$ terkecil.
3. Menentukan baris kunci (variabel keluar) dengan memilih nilai rasio terkecil antara nilai ruas kiri (b_i) dengan koefisien kolom kunci (a_{ij}). Perhitungan rasio dilakukan dengan $Rasio = \frac{b_i}{a_{ij}}$, di mana $rasio > 0$.
4. Menentukan pivot pada elemen kunci yang terletak pada perpotongan antara kolom kunci dan baris kunci dalam algoritma simpleks. Elemen kunci ini kemudian diubah nilainya menjadi 1.
5. Melakukan Operasi Baris Dasar (OBD) berdasarkan pada pivot untuk baris-baris lainnya, termasuk $c_j - z_j$. Elemen-elemen dalam kolom kunci pada baris ini dijadikan nol, kecuali elemen yang dijadikan pivot. $Baris\ kunci\ baru = \frac{Baris\ kunci\ lama}{pivot}$
   $Baris\ baru\ selain\ baris\ kunci = Baris\ lama - Unsur\ kolom\ kunci \times Baris\ kunci\ baru$
6. Proses iterasi untuk masalah maksimum akan berhenti jika semua nilai pada baris $c_j - z_j \leq 0$ terpenuhi, menandakan solusi telah optimal. Apabila masih terdapat nilai $c_j - z_j > 0$ (positif), maka iterasi algoritma simpleks perlu dilanjutkan. Adapun untuk masalah minimum, proses iterasi akan berhenti jika semua nilai pada baris $c_j - z_j \geq 0$ terpenuhi. Apabila masih terdapat nilai $c_j - z_j < 0$ (negatif), maka iterasi algoritma simpleks perlu dilanjutkan (Khairunisa, 2015).

1.5.4 Regresi *Spline*

Spline adalah model regresi nonparametrik yang terdiri dari potongan fungsi polinomial yang memiliki sifat tersegmen dan kontinu. Fungsi polinomial ini memiliki karakteristik khusus yang memberikan fleksibilitas lebih tinggi dibandingkan dengan fungsi polinomial pada umumnya. Selain itu, dalam fungsi ini terdapat titik-titik penghubung yang disebut dengan titik knot. Titik knot adalah perpaduan kurva yang menunjukkan pola perubahan perilaku pada selang yang berbeda (Nurdiani dkk., 2017). *Spline* memiliki keunggulan dalam mengatasi pola data yang menunjukkan kenaikan atau penurunan dengan bantuan titik-titik knot serta kurva yang dihasilkan relatif mulus (Utami, 2018). Secara umum, fungsi *spline* berorde q dengan titik knot pada $K_1, K_2, \dots, K_r$ adalah sebagai berikut (Putra dkk., 2015).

$$f(x_i) = \sum_{l=0}^{q} \beta_l x_i^l + \sum_{h=1}^{r} \beta_{q+h}(x_i - K_h)_+^q \tag{15}$$

di mana fungsi *polynomial truncated* pada Persamaan (15) adalah sebagai berikut.

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)^q, & x \geq K_h \\ 0 & , x < K_h \end{cases} \tag{16}$$

dengan

$f(x_i)$	:	fungsi regresi
β_l	:	parameter polinomial pada orde ke-l
x_i	:	variabel prediktor pada pengamatan ke-i
$\beta_0, \beta_1, \dots, \beta_{q+h}$	:	parameter *truncated*
K_h	:	titik knot ke-h, $(h = 1,2, \dots, r)$
$(x_i - K_h)_+^q$	:	fungsi *polynomial truncated*

Model yang menyatakan antara k variabel prediktor dengan satu variabel respon, jika dinyatakan dalam bentuk fungsi *spline* $f(x_i)$ dalam Persamaan (15), maka dapat dinyatakan sebagai berikut (Pratiwi, 2020).

$$f(x_{ij}) = \sum_{l=0}^{q} \beta_{lj} x_{ij}^l + \sum_{h=1}^{r} \beta_{(q+h)j}\left(x_{ij} - K_{hj}\right)_+^q \tag{17}$$

dengan $i = 1,2, \dots, n$ dan $j = 1,2, \dots, k$

$f(x_{ij})$	:	fungsi regresi *spline* dari pengamatan ke-i pada variabel prediktor ke-j
x_{ij}	:	pengamatan ke-i pada variabel prediktor ke-j
β_{lj}	:	parameter polinomial pada orde ke-l dan prediktor ke-j
$\beta_{(q+h)j}$	:	parameter *truncated* pada titik knot ke-$(q+h)$ dan prediktor ke-j
K_{hj}	:	nilai titik knot ke-h pada prediktor ke-j
ε_i	:	*error* pada pengamatan ke-i

Apabila model pada Persamaan (17) disajikan ke dalam bentuk matriks, maka dapat dituliskan sebagai berikut.

$$\boldsymbol{y} = \boldsymbol{X}[K]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{18}$$

di mana $\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]'$ adalah vektor berukuran $n \times 1$, $\boldsymbol{X}[K] = [\boldsymbol{1} \quad \boldsymbol{X_1} \quad \dots \quad \boldsymbol{X_k}]$ adalah matriks $\boldsymbol{X}$ dalam bentuk *spline* dengan orde q dan r titik knot berukuran $n \times (1 + (q + r) \times k)$ dengan

$$1 = [1 \quad 1 \quad \cdots \quad 1]', \boldsymbol{X_1} = \begin{bmatrix} x_{11} & x_{11}^2 & \cdots & x_{11}^q & (x_{11} - K_{11})_+^q & \cdots & (x_{11} - K_{r1})_+^q \\ x_{21} & x_{21}^2 & \cdots & x_{21}^q & (x_{21} - K_{11})_+^q & \cdots & (x_{21} - K_{r1})_+^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n1}^2 & \cdots & x_{n1}^q & (x_{n1} - K_{11})_+^q & \cdots & (x_{n1} - K_{r1})_+^q \end{bmatrix},$$

$$\boldsymbol{X_k} = \begin{bmatrix} x_{1k} & x_{1k}^2 & \cdots & x_{1k}^q & (x_{1k}-K_{1k})_+^q & \cdots & (x_{1k}-K_{rk})_+^q \\ x_{2k} & x_{2k}^2 & \cdots & x_{2k}^q & (x_{2k}-K_{1k})_+^q & \cdots & (x_{2k}-K_{rk})_+^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{nk} & x_{nk}^2 & \cdots & x_{nk}^q & (x_{nk}-K_{1k})_+^q & \cdots & (x_{nk}-K_{rk})_+^q \end{bmatrix}$$

$\boldsymbol{\beta} = [\beta_0 \quad \beta_{11} \quad \cdots \quad \beta_{q1} \quad \beta_{(q+1)1} \quad \cdots \quad \beta_{(q+r)1} \quad \cdots \quad \beta_{1k} \quad \cdots \quad \beta_{qk} \quad \beta_{(q+1)k} \quad \cdots \quad \beta_{(q+r)k}]'$ adalah vektor berukuran $(1+(q+r)\times k)\times 1$, dan $\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]'$ adalah vektor berukuran $n \times 1$ (Putra dkk., 2015).

1.5.5 Estimasi Fungsi Regresi *Spline*

(Hidayat dkk., 2017) melakukan sebuah penelitian menggunakan kurva regresi f yang dijelaskan menggunakan fungsi *spline* f titik knot K. Bentuk matriks dari fungsi tersebut disajikan sebagai berikut.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{19}$$

Kemudian, estimasi parameter $\boldsymbol{\beta} = \left[\beta_0, \beta_1, \dots, \beta_q, \dots, \beta_{q+r}\right]'$ diperoleh melalui metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat *error* sebagai berikut.

$$\min_{\beta \in \mathbb{R}^{q+1+h}} \{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\} = \min_{\beta \in \mathbb{R}^{q+1+h}} \left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}' \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \right)$$

$$\min_{\beta \in \mathbb{R}^{q+1+h}} \{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\} = \min_{\beta \in \mathbb{R}^{q+1+h}} \{[\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta}'][[\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta}]\}$$

dengan penyajian matriks diberikan sebagai berikut.

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^2 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= ([\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta}'][[\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta}]) \\ &= ([\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}]'[\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}]) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta} \end{aligned} \tag{20}$$

Kemudian, Persamaan (20) diturunkan terhadap vektor $\boldsymbol{\beta}'$ dan disamakan dengan nol, maka diperoleh hasil sebagai berikut.

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = \frac{\partial(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'[K]\boldsymbol{X}[K]\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \tag{21}$$

dengan

$$\begin{aligned} \boldsymbol{X}[K] &= \boldsymbol{X}[K_1, K_2, \dots, K_r] \\ &= \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^q & (x_1-K_1)_+^q & \cdots & (x_1-K_r)_+^q \\ 1 & x_2 & x_2^2 & \cdots & x_2^q & (x_2-K_1)_+^q & \cdots & (x_2-K_r)_+^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^q & (x_n-K_1)_+^q & \cdots & (x_n-K_r)_+^q \end{bmatrix} \end{aligned}$$

Selanjutnya, diperoleh estimasi fungsi regresi *spline* sebagai berikut.

$$\hat{\boldsymbol{f}}[K](x_i) = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]\boldsymbol{y} \quad (22)$$

Persamaan (21) dapat dinyatakan ke dalam bentuk sebagai berikut.

$$\hat{\boldsymbol{f}}[K](x_i) = \hat{\boldsymbol{y}} = \boldsymbol{A}[K]\boldsymbol{y} \quad (23)$$

dengan $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]$ dan $\boldsymbol{X}[K]$ adalah matriks dari model yang bergantung pada titik knot dan $K = [K_1, K_2, \dots, K_r]'$ adalah titik knot. Dapat dilihat bahwa $\hat{\boldsymbol{f}}[K](x_i)$ adalah estimator dalam observasi $\boldsymbol{y} = [y_{\mathbf{1}}, y_2, \dots, y_n]'$.

1.5.6 Pemilihan Titik Knot Optimal

Salah satu tahapan penting dalam pendekatan *spline* adalah pemilihan titik knot yang optimal. Titik knot adalah titik perpaduan bersama, di mana terjadi perubahan dalam perilaku fungsi di interval yang berbeda. Pemilihan titik knot optimal diperlukan untuk menentukan model *spline* yang optimal. Salah satu metode pemilihan titik knot optimal adalah dengan menggunakan *Generelized Cross Validation* (GCV) minimum (Marina & Budiantara, 2013). Titik knot optimal untuk model *spline* terbaik diperoleh dari nilai GCV yang terkecil. Metode GCV dapat dituliskan sebagai berikut.

$$GCV(K) = \frac{MSE(K)}{(n^{-1}trace[\boldsymbol{I} - \boldsymbol{A}(K)])^2} \quad (24)$$

dengan $K = [K_{11}, K_{12}, \dots, K_{1h}, \dots, K_{p1}, \dots, K_{pr}]$ adalah titik knot, $MSE(K) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, dan $\boldsymbol{A}[K] = \boldsymbol{X}[K](\boldsymbol{X}'[K]\boldsymbol{X}[K])^{-1}\boldsymbol{X}'[K]$ (Meimela, 2020).

1.5.7 Demam Berdarah *Dengue*

Demam Berdarah *Dengue* (DBD) adalah suatu penyakit infeksi yang disebabkan oleh virus *dengue*. Gejalanya meliputi demam, nyeri otot atau nyeri sendi, *leukopenia*, ruam, *limfadenopati*, *trombositopenia*, dan *diastes hemorganik*. DBD ditandai oleh perembesan plasma yang mengakibatkan *hemokonsentrasi* atau penumpukan cairan di dalam tubuh. Virus *dengue* masuk ke tubuh manusia melalui gigitan nyamuk *Ae Aegypti* atau *Ae Albopictus*. Virus ini menyerang organ-organ, seperti *hepar*, *nodus limfaticus*, sumsum tulang, dan paru-paru. Setelah nyamuk menggigit individu yang terinfeksi, virus berkembang biak dalam kelenjer liur selama delapan hingga sepuluh hari sebelum dapat ditularkan kembali kepada manusia melalui gigitan berikutnya. Infeksi pertama mungkin hanya menimbulkan gejala demam *dengue*, sementara DBD biasanya terjadi setelah seseorang mengalami infeksi berulang akibat virus *dengue* yang berbeda, yang dikenal sebagai infeksi *the secondary heterologous infection* atau *the sequential infection hypothesis* (Hadijah, 2020).

DBD dapat berkembang menjadi *Syndrome Syok Dengue* (SSD), yang dapat berakibat fatal tanpa penanganan yang tepat. Penting untuk mendiagnosis DBD dengan akurat dan memantau perkembangannya secara cermat. Pada awal DBD, jumlah leukosit dapat menurun, tetapi kemudian dapat kembali normal dengan dominasi sel neutrofil. Ketika penyakit memasuki tahap akhir, jumlah total leukosit dan sel *polimorfonuklear* akan menurun, sementara limfosit akan meningkat, sehingga menyebabkan pergeseran ke kanan dalam perhitungan jenis leukosit (*shift to the right*). Leukosit atau yang sering disebut sel darah putih merupakan komponen penting dalam sistem kekebalan tubuh yang bertanggung jawab untuk melawan infeksi dari virus, bakteri, dan toksin metabolik. Ada lima jenis leukosit, masing-masing memiliki peran khusus dalam melawan patogen, yaitu neutrofil, eosinofil, basofil, monosit, dan limfosit. Hasil perhitungan jenis leukosit sangat membantu dalam diagnosis yang akurat dan memberikan informasi yang spesifik tentang infeksi dan perkembangan penyakit (Purwanto, 2002).

Kriteria laboratorium *World Health Organization* (WHO) yang menunjukkan jumlah trombosit yang rendah (*trombositopenia*) dan kebocoran plasma yang ditandai dengan *hemokonsentrasi* adalah indikator keparahan penyakit DBD. *Trombositopenia* disebabkan oleh penurunan produksi trombosit di sumsum tulang, sementara mekanisme yang mengakibatkan peningkatan penggunaan trombosit memainkan peran utama dalam menginduksi *trombositopenia* pada DBD. Penurunan jumlah trombosit biasanya terjadi sebelum terdapat peningkatan hematokrit dan sebelum suhu tubuh turun. *Trombositopenia* dianggap terjadi ketika jumlah trombosit turun di bawah 10^3 sel/μl, biasanya terjadi di antara hari ketiga hingga ketujuh setelah sakit (Hadijah, 2020).

BAB II
METODE PENELITIAN

2.1 Sumber Data dan Variabel

Penelitian ini menggunakan data sekunder yang diperoleh dari hasil rekam medis pasien penderita DBD di Rumah Sakit Pendidikan Unhas selama tahun 2013-2017. Variabel penelitian terdiri dari variabel respon dan variabel prediktor. Variabel respon dalam penelitian ini adalah kadar trombosit pada pasien penderita DBD yang diukur melalui hasil uji laboratorium pada hari pertama mereka dirawat di rumah sakit dengan kadar rata-rata sebesar 10^3 sel/μl. Selanjutnya, variabel prediktor yang digunakan dalam penelitian ini terdiri sebagai berikut.

1. Suhu tubuh (x_1), merujuk pada suhu tubuh pasien yang diukur pada hari pertama pasien dirawat inap di rumah sakit yang dinyatakan dalam °C.
2. Sel darah putih (x_2), merujuk pada hasil uji laboratorium pasien yang diukur pada hari pertama pasien dirawat inap di rumah sakit yang dinyatakan dalam 10^3 sel/μl.

2.2 Metode Analisis

Pendekatan analisis yang digunakan dalam penelitian ini menggunakan model regresi kuantil dengan estimator *spline* kubik multivariabel. Adapun prosedur analisis yang digunakan untuk mencapai tujuan penelitian adalah sebagai berikut.

2.2.1 Estimasi Model Regresi Kuantil dengan Estimator *Spline* Kubik Multivariabel

1. Diberikan data $\{x_{i1}, x_{i2}, \dots, x_{ik}, y_i\}$, dengan $i = 1,2,\dots,n$ dan $j = 1,2..,k$, yang merupakan himpunan berpasangan dari variabel acak yang berdistribusi secara independen dan tidak identik serta mengikuti model regresi kuantil linier dengan kuantil $\tau \in (0,1)$.
$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_j(\tau)x_{ij} + \cdots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau)$$
2. Menyajikan model regresi kuantil dengan estimator *spline* kubik multivariabel.
$$y_i(\tau) = \sum_{j=1}^{k} f(x_{ij}) + \varepsilon_i(\tau)$$
3. Menyajikan model regresi kuantil dengan estimator *spline* kubik multivariabel dalam bentuk matriks.
$$\boldsymbol{y}(\tau) = \boldsymbol{X}[K]\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau)$$
4. Memperoleh estimasi parameter $\boldsymbol{\beta}$ dengan meminimumkan persamaan berikut menggunakan algoritma simpleks.
$$\hat{\beta}(\tau) = \frac{min}{\beta(\tau)}\left\{\tau\sum_{i=1}^{n}|\varepsilon_{i1}(\tau)| + (1-\tau)\sum_{i=1}^{n}|\varepsilon_{i2}(\tau)|\right\}$$

5. Memperoleh estimasi model regresi kuantil dengan estimator *spline* kubik multivariabel.

$$\widehat{\boldsymbol{y}}(\tau) = \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}(\tau)$$

2.2.2 Pemodelan Regresi Kuantil dengan Estimator *Spline* Kubik Multivariabel

1. Melakukan eksplorasi data dari variabel respon, yaitu kadar trombosit dan variabel prediktor, yaitu suhu tubuh dan sel darah putih untuk mengetahui statistik deskriptif dari masing-masing variabel penelitian.
2. Melakukan spesifikasi model menggunakan *scatter plot* untuk mengetahui bentuk pola hubungan dari variabel respon terhadap masing-masing variabel prediktor.
3. Mendeteksi pencilan pada data.
4. Memodelkan data menggunakan regresi kuantil dengan estimator *spline* kubik multivariabel dengan nilai kuantil, yaitu $\tau = 0.25,\ 0.50$, dan 0.75 pada satu hingga tiga titik knot.
5. Memperoleh model regresi kuantil dengan estimator *spline* kubik multivariabel yang optimal melalui nilai GCV minimum.
6. Menginterpretasikan kadar trombosit berdasarkan suhu tubuh dan sel darah putih pada pasien penderita DBD yang diperoleh menggunakan model regresi kuantil dengan estimator *spline* kubik multivariabel.